A325-1023-7M-L 180

TEXAS AGRICULTURAL EXPERIMENT STATION

AGRICULTURAL AND MECHANICAL COLLEGE OF TEXAS

W. B. BIZZELL, President

]LLETIN NO. 310 SEPTEMBER, 1923

DIVISION OF AGRONOMY

THE INTERPRETATION OF CORRELATION DATA

LIBRARY
TNAU, Coimbatore - 3
000037565

B. YOUNGBLOOD, DIRECTOR

COLLEGE STATION, BRAZOS COUNTY, TEXAS.

STATION STAFF†

ADMINISTRATION
B. Youngblood, M. S., Ph. D., *Director*
A. B. Conner, M. S., *Vice Director*
A. H. Leidigh, M. S., *Assistant Director*
C. A. Felker, *Chief Clerk*
A. S. Ware, *Secretary*
M. P. Holleman, Jr., *Assistant Chief Clerk*
J. M. Schaedel, *Executive Assistant*
C. J. Gorzycki, *Technical Assistant*

VETERINARY SCIENCE
*M. Francis, D. V. M., *Chief*
H. Schmidt, D. V. S., *Veterinarian*
V. J. Brauner, D. V. M., *Veterinarian*

CHEMISTRY
G. S. Fraps, Ph. D., *Chief; State Chemist*
S. E. Asbury, M. S., *Assistant Chemist*
W. H. Walker, *Assistant Chemist*
A. G. Peterson, B. S., *Assistant Chemist*
J. E. Teague, B. S., *Assistant Chemist*
J. K. Blum, B. S., *Assistant Chemist*

HORTICULTURE
A. T. Potts, M. S., *Chief*

ANIMAL INDUSTRY
J. M. Jones, A. M., *Chief*
J. L. Lush, Ph. D., *Animal Husbandman, Breeding*
G. R. Warren, B. S., *Swine Husbandman*
R. M. Sherwood, B. S., *Poultry Husbandman*
J. J. Hunt, *Wool Grader*

ENTOMOLOGY
M. C. Tanquary, Ph. D., *Chief; State Entomologist*
H. J. Reinhard, B. S., *Entomologist*
C. S. Rude, B. S., *Entomologist*
W. O. Victor, Jr., *Apiary Inspector*
W. R. Jordan, B. S., *Apiary Inspector*
H. J. Morris, *Apiary Inspector*

AGRONOMY
E. B. Reynolds, M. S., *Chief*
G. N. Stroman, Ph. D., *Cotton Breeder*

PLANT PATHOLOGY AND PHYSIOLOGY
J. J. Taubenhaus, Ph. D., *Chief*

FARM AND RANCH ECONOMICS
L. P. Gabbard, M. S., *Chief*
V. L. Cory, M. S., *Grazing Research Botanist*
H. E. Rea, B. S., *Assistant*

SOIL SURVEY
**W. T. Carter, Jr., B. S., *Chief*
H. W. Hawker, *Soil Surveyor*
Edward Templin, *Soil Surveyor*

BOTANY
H. Ness, M. S., *Chief*

PUBLICATIONS
A. D. Jackson, *Chief*
C. M. Mitchell, *Mailing Clerk*

STATE APICULTURAL RESEARCH LABORATORY
H. B. Parks, B. S., *Apiculturist in Charge*
A. H. Alex, B. S., *Queen Breeder*

MAIN STATION FARM
D. T. Killough, B. S., *Superintendent*

FEED CONTROL SERVICE
B. Youngblood, M. S., Ph. D., *Director*
F. D. Fuller, M. S., *Chief*
S. D. Pearce, *Secretary*
J. H. Rogers, *Inspector*
W. H. Wood, *Inspector*
J. J. Kelly, *Inspector*
J. D. Prewit, B. S., *Inspector*
D. W. Carlton, B. S., *Inspector*

SUBSTATIONS

No. 1. Beeville, Bee County
I. E. Cowart, M. S., *Superintendent*

No. 2. Troup, Smith County
W. S. Hotchkiss, *Superintendent*

No. 3. Angleton, Brazoria County
V. E. Hafner, B. S., *Superintendent*

No. 4. Beaumont, Jefferson County
A. H. Prince, B. S., *Superintendent*

No. 5. Temple, Bell County
A. B. Cron, B. S., *Superintendent*

No. 6. Denton, Denton County
P. B. Dunkle, B. S., *Superintendent*

No. 7. Spur, Dickens County
R. E. Dickson, B. S., *Superintendent*

No. 8. Lubbock, Lubbock County
R. E. Karper, B. S., *Superintendent*

No. 9. Balmorhea, Reeves County
J. J. Bayles, B. S., *Superintendent*

No. 10. College Station, Brazos County (Feeding and Breeding Substation)
L. J. McCall, *Superintendent*

No. 11. Nacogdoches, Nacogdoches County
G. T. McNess, *Superintendent*

****No. 12. Chillicothe, Hardeman County**
D. L. Jones, *Acting Superintendent*

No. 14. Sonora, Sutton-Edwards Counties
E. M. Peters, B. S., *Superintendent*
D. H. Bennett, V. M. D., *Veterinarian*
O. L. Carpenter, B. S., *Shepherd*

Teachers in the School of Agriculture Carrying Cooperative Projects on the Station
‡S. W. Bilsing, *Professor of Entomology*
W. L. Stangel, *Professor of Animal Husbandry, Hogs*
F. A. Buechel, *Professor of Agricultural Economics.*
G. W. Adriance, *Associate Professor of Horticulture*
W. E. Garnett, *Professor of Rural Sociology*
G. P. Grout, *Professor of Dairy Husbandry*

---

†As of October 1, 1923.
*In cooperation with School of Veterinary Medicine, A. & M. College of Texas.
**In cooperation with United States Department of Agriculture.
†On leave.

CONTENTS.

	PAGE
Summary	5
Introduction	7
The Kinds of Correlation	9
The Correlation Table and the Placement of Material	10
The Interpretation of Population Data	13
The Interpretation of Pure-Line Data	19
The Comparative Interpretation of Population and Pure-Line Data	21

SUMMARY

1. Correlation implies a causal relationship or connection. Correlation data dealing with biological material should have the most careful analysis, giving due consideration to the causal agency or agencies if the data are to be of the greatest value.

2. To summarize the different kinds of correlation: the most workable classification for the student of biological data to use is one grouping the kinds of correlation into two classes—genetic and non-genetic.

3. The general features of the correlation table are shown and illustrations given of the meaning of different placements of frequency distributions in the table, and showing also the possibilities in the analysis of material where the non-genetic correlation is known.

4. In interpreting population data one should bear in mind that the correlation obtained may apply only to the particular population studied. It is evident that population material may be of doubtful value in drawing conclusions as to correlation between characters. Accordingly, one should be very cautious not to overestimate its value.

5. In the interpretation of pure-line data one can determine with certainty the measure of the non-genetic correlation. From pure-line correlation data one may also determine the variation in non-genetic correlations from season to season. From line data one may show by correlation whether genetic variation exists and hence determine the stability of the line with respect to the characters studied.

6. Since the pure line gives a reliable measure of non-genetic influences, the comparative use of pure-line and population data will permit in many cases the determination with certainty that genetic correlation exists, and with this knowledge one may with certainty isolate families showing such correlation and by other means determine the nature of the genetic correlation.

Bulletin No. 310 September, 1923

THE INTERPRETATION OF CORRELATION DATA*

A. B. CONNER

Correlation, in so far as it relates to biological data, is the relation that exists between characters due to a common causal agency or agencies. It implies a causal relationship or connection and without such connection there can be no correlation. King[1] illustrates the lack of correlation that would exist between the cocoanut crop of the Fiji Islands and the money supply of the United States, unless it could be shown that one was the cause of the other or that both changes were due to some common factor.

Davenport[2] states: "The whole subject of correlation refers to that inter-relation between separate characters by which they tend, in some degree at least, to move together. This relation is expressed in the form of a ratio."

The underlying fact that there must be a causal relation between the characters under consideration is, it seems, the chief reason why the correct interpretation of correlation data is difficult. The fact that correlation exists can be established, however, with certainty without absolute knowledge of the nature of the influence. Nevertheless, any interpretation of data must give due consideration to the operations of the causal influences.

Collins[3] states that correlation studies were at one time thought to be full of promise as an aid to the plant breeder, but that in recent years little use has been made of correlation by practical breeders. He says: "Yet it must be admitted on reflection that nearly all successful breeding has in reality been made possible by the fact that correlations exist. * * * The existence of types must mean that there are many individuals that present approximately the same combination of characters, and this is exactly what correlation implies. * * * If the study of correlations has appeared to have little bearing on plant breeding, it must be that we have been studying the wrong characters or studying them in the wrong way."

Babcock and Clausen[4] state: "A fine illustration of what the bio-

*Submitted to the Faculty of the Agricultural and Mechanical College of Texas, in June, 1923, in partial fulfillment of the requirements for the degree of Master of Science in Agriculture.

[1]King, W. I., "The Elements of Statistical Method," the Macmillan Company, New York, 1921, page 197.

[2]Davenport, E., "Principles of Breeding," Ginn and Company, New York, 1907, page 453.

[3]Collins, G. N., "Correlated Characters in Maize Breeding," Journal of Agricultural Research, Vol. 6, No. 12, June 19, 1916, page 435.

[4]Babcock and Clausen, "Genetics in Relation to Agriculture," pp. 506-7.

metrician can do in this line is found in the recent work on the correlation between body pigmentation and egg production in the domestic fowl, by Harris, Blakeslee and Warner. This study dealt with the relationship between the concentration of yellow pigment in the ear lobe of White Leghorn hens and their egg records of the preceding month. It was found that there is a very close interdependence between October ear lobe color and the egg production of the pullet year.' They[5] further state: "The mathematical relations existing in linkage phenomena are of interest because they provide a method of determining the genetic relationships involved in certain cases of somatic correlations. If two factors are linked in inheritance, it follows that a larger proportion of the population will display the corresponding two characters than would be the case if the factors were inherited independently. Consequently, character correlations of this type are an index to factor linkage."

Babcock and Clausen[6] further refer to the use of the coefficient of correlation as a measure of intensity of inheritance as a practice of doubtful scientific propriety and one which might favor misleading conclusions. The lack of correlation between parent and offspring may be the case as a consequence of genetic variability. On the other hand "modifiability" may be a factor determining the value of the correlation

Babcock and Clausen[7] state that Pearl and Surface show a lack of correlation between mothers and daughters in egg production when mass selection was practiced. They subsequently show, however, that marked increases in egg production were obtained by selection of genotypes. This indicates, as pointed out by Babcock and Clausen, that some method of breeding must be adopted that will discount at their proper values the influences of modifiability and genetic variability attending segregation.

Hayes and Garber[8] discuss the early work in correlation of plant characters and yield, particularly that of the Svalof Station with oats, and the Minnesota and Nebraska Stations with wheat. They further point out that though the earlier work indicated correlation between certain characters and yield, subsequent pure-line work showed that in general no one character is closely associated with yield, at least to such an extent as to be of selection value in picking out high yielding strains. It may be said, however, that yield is the result of many growth factors and that many of the characters studied could hardly be classed as being expressions of growth factors.

It is very evident that correlation data dealing with biological material should have the most careful analysis, and in order to do this the nature of the material used must be known and full consideration given to the uses which can be made of it if the data are to be of the greatest value. It should be borne in mind that correlation data are essentially descriptive of the material in hand. It is the purpose of

[5]Babcock and Clausen, "Genetics in Relation to Agriculture, page 127.

[6]Ibid., page 459.

[7]Ibid., pages 457-458.

[8]Hayes, H. K., and Garber, R. J., "Breeding Crop Plants," First Edition, McGraw-Hill Book Company, Inc., New York, 1921, pages 125, 126, 127.

this paper to discuss the interpretation of correlation data in a way that will give the student a clear perspective of the uses and limitations of material of this character.

THE KINDS OF CORRELATION

Webber[9] distinguishes four kinds of correlation which he says should be recognized. These are termed environmental, morphological, physiological, and coherital. He describes the environmental correlations as expressions of physical conditions due to varying conditions of fertility or other environmental causes. He cites Leibenberg's work in 1892 and 1893, showing that the length of stem in wheat is correlated with increase in the strength of stem, the length of head, the number of spikelets, the number of kernels, and the total weight of kernels produced; and he further refers to similar observations of Proskowetz in barley and of Fruwirth in field beans, in which such correlations were considered as merely the expression of a condition of luxuriance. He adds that, strictly speaking, these are not correlated characters and their consideration is of little or no value to the breeder. He describes morphological correlations as those cases wherein a relation of one character is the primary cause for the variation in another character. He states that this type of correlation, which at first seems similar to coherital correlation, is, he thinks, of an entirely different nature, since the two characters are intimately related in a morphological or a physiological sense and increase in one organ necessarily gives rise to an increase in the other. Webber's physiological correlations are illustrated by the relation of the number of leaves to seed production in tobacco, the heavy leaf production being correlated with lack of seed production for the reason that the main strength of the plant goes into the leaves at the expense of seed production.

Webber's fourth group of correlations is described by him as coherital and include those characters which are not related to each other in any direct or causal sense, but which are inherited as a single unit character. Linked characters would undoubtedly fall into this class. He tates that correlations belonging to this group are the most interestng ones from a scientific standpoint and in some cases may be of great ractical value. He cites his work in hybridizing corn, in which he tates that certain characters hang together in the splitting up of the ybrids instead of the expected breaking down of the correlation which ohannsen states is the result of crossing. Webber notes, however, hat in about one case out of fifty or one hundred the correlation is roken. He is evidently referring here to crossing-over.

East[10] refers to Webber's classification of correlations and states hat considering all the types of correlation, without regard to whether r not they are of value to the breeder, they fall naturally into two lasses—somatic and gametic. Undoubtedly his classification of cor-

[9]Webber, H. J., "Correlation of Characters in Plant Breeding," American reeders Association Report, Volume 2, 1906, pages 73, 74, 75.

[10]East, E. M., "Organic Correlations," American Breeders Association Report, olume 4, 1908, page 333.

relations into two groups will greatly simplify the interpretation of correlation data, as, after all, it is of little importance to distinguish as between correlations of different groups except in so far as they are internal or external, or, in other words, inherited or not inherited.

Collins[11] classifies correlations as physical, physiological, and genetic. His classification of physical correlations is referred to as those such as exist when increased weight is correlated with increased height, and he further states that this kind of correlation would be found in stones and inanimate objects selected at random. Collins describes his physiological correlations as being the result of the same physiological tendency. He states that "Genetic correlations cover the large residue of correlations, the nature and causes of which are questions of controversy, but which are associated with the method or mechanism of heredity." It would seem that perhaps Collins' physical correlations might well be classified as genetic, since it is a debatable question as to whether correlation of increased weight with increased height in a plant is a similar relationship to that which may be found in stones or inanimate objects selected at random.

To sum up the different classifications of correlations: it seems that the most workable classification that can be made has been suggested by East,[12] in which he classes all correlations as somatic or gametic. At any rate, in analyzing data we are concerned primarily with the genetic and the non-genetic relationships between the characters involved.

THE CORRELATION TABLE AND THE PLACEMENT OF MATERIAL

The correlation table represents the frequency distributions for the two different characters being studied. Babcock and Clausen[13] present the general features of a correlation table in the following diagram, in which V equals the variation of any individual from the mean and M equals the mean:

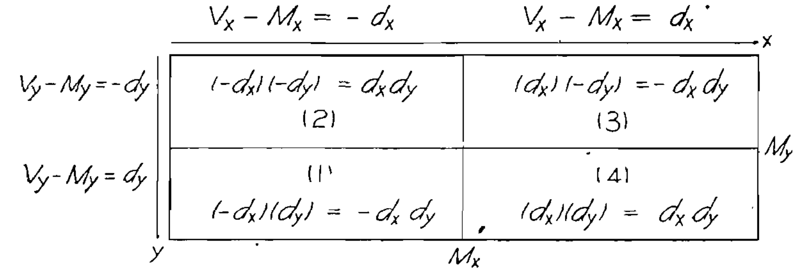

Figure 1.

From Babcock and Clausen's Genetics in Relation to Agriculture
By permission of the McGraw-Hill Book Company, Inc., New York.

[11]Collins, G. N., "Correlated Characters in Maize Breeding," Journal of Agricultural Research, Volume 6, No. 12, June 19, 1916, pages 436-37.

[12]East, E. M., "Organic Correlations," American Breeders Association Report, Volume 4, 1908, page 333.

[13]Babcock and Clausen, "Genetics in Relation to Agriculture," page 51, Fig. 24.

This diagram is based upon the intersection of the two means dividing the table into four parts. Accordingly, distributions that fall in greatest numbers in quadrants 2 and 4 would represent a positive correlation. In other words, as one of the characters is increased the other increases. Distributions that fall in greatest numbers in quadrants 1 and 3 represent negative correlation, that is, as one character increases the other decreases. Such placements, as well as a placement where no correlation would exist, are shown in the following figures taken from Babcock and Clausen:[14]

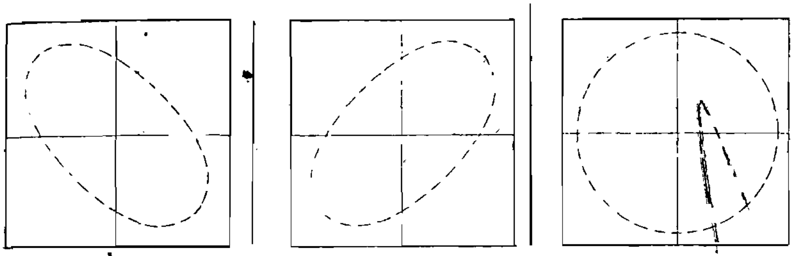

Figure 2.
From Babcock and Clausen's Genetics in Relation to Agriculture
By permission of the McGraw-Hill Book Company, Inc., New York

Along with these simple illustrations of placement of material in a correlation table, it is well to present in a general way the manner in which causal agencies might affect this placement and, hence, the kind and degree of correlation. To illustrate the manner in which placement is affected, let us take the classification of all correlations as being genetic or non-genetic and then show the placement of pure-line material where all the individuals considered are supposed to have the same genetic constitution. The placement, therefore, with respect to the two characters being studied would be an arrangement according to the manner in which non-genetic factors had caused the two characters to vary in this material. The placement here would be positive, negative, or neutral. In any event, it represents the placement when non-genetic factors only have influenced these two characters. We have here, it seems, in the correlation table made up of pure-line material, a measure of the influence of non-genetic factors. Similarly, correlation tables based on other pure-line material for the same character would show similar influences unless different characters can be shown to react in contrary directions to the first case. Accordingly, we may conclude that if we have correlation coefficients on a number of pure lines they will consistently show similar correlations, and these furnish a reliable measure of non-genetic influences.

Let us take a population composed of two pure lines which differ widely with respect to hereditary composition for the two characters, and construct a correlation table. While such material represents an extreme case, it, nevertheless, serves to illustrate the placement of population material. The result would be the placement of material in one of the four ways illustrated in the following diagrams:

[14]Babcock and Clausen, "Genetics in Relation to Agriculture," page 52, Fig. 25.

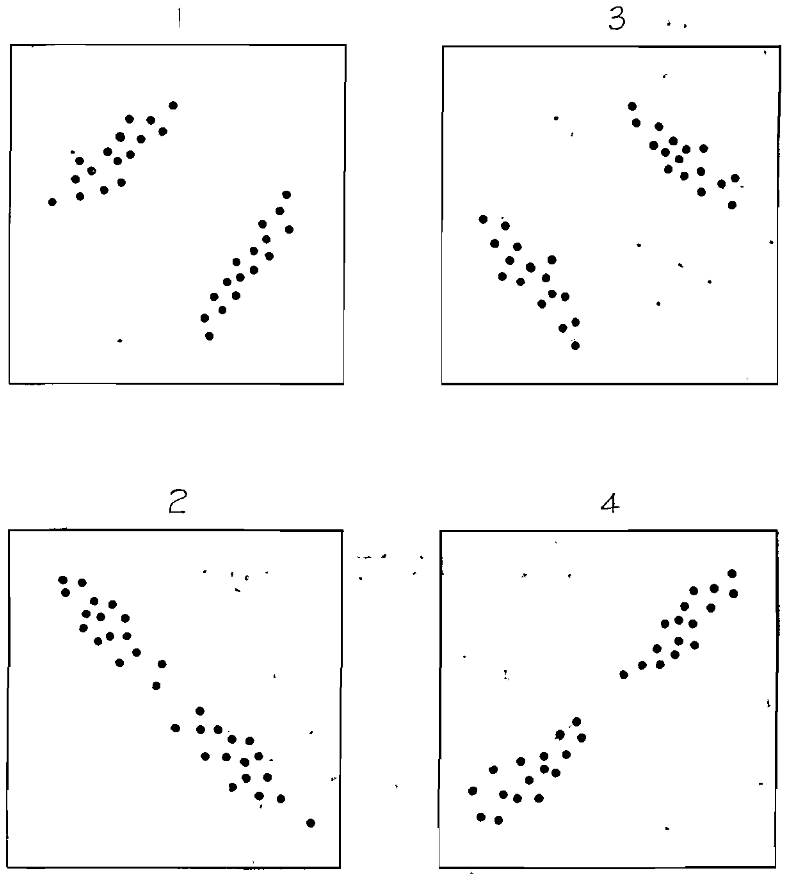

FIGURE 3

Diagrams 1 and 2 would be the possible arrangement or placement of the material if the genetic correlation existing in this population were positive, while Diagrams 3 and 4 would represent its possible placement if the genetic correlation were negative. Diagrams 1 and 4 show a placement of the lines in the table within the population as influenced by non-genetic factors in a negative way, whereas Diagrams 2 and 3 show the lines if influenced in a positive way. In the event that non-genetic correlations between two characters are negative, the placement of correlation data in population material in a positive way would indicate the certainty of genetic differences of some of the families composing the population. It is realized that in many instances genetic differences would be obscured by the average placement of families composing the population, but, nevertheless, wide genetic differences are frequently so indicated by exaggeration or a tendency toward reversal of the correlation.

THE INTERPRETATION OF POPULATION DATA

In the interpretation of population data one should bear in mind that he has a series of distributions within the table whose placement may have been influenced either by genetic differences of families comprising the population or by non-genetic influences, such as favorable or unfavorable environment. The material in hand may be influenced in the same or in opposite directions by these genetic and non-genetic influences. It is generally unknown how many families have been included in the population, and, moreover, the genetic relationships of these families are likewise unknown. Accordingly, we have a mass of material in the form of a correlation table in which the placements may perhaps have been influenced by both genetic and non-genetic factors, from which we expect to glean some information as to the correlation of characters. The general trend of the material will likely show whether or not correlation exists in this population and the extent of such correlation, if it exists; however, it is not enough to know that the particular material in hand shows correlation, for the whole purpose of the interpretation is to gain knowledge as to whether or not two characters have a more or less consistent relation to each other. One may determine this relationship with certainty from population material, provided he has correlation tables for the same characters on a number of different populations. The consistent revelation of correlation between two characters in population material may be the result of linkage or other association of characters, in which case the characters will be found associated more often than otherwise. On the other hand, the existence of correlation between two characters in different population material might be the result of non-genetic influences. For example, in kafir grown at Substation No. 8, Lubbock, Texas, different lines have shown correlations between weight of head and diameter of plant, as follows:

Line 153: $r = .5501 \pm .0413$
Line 567: $r = .6254 \pm .0336$
Line 192: $r = .7509 \pm .0233$
Line 40: $r = .5474 \pm .0428$

Now, since within any one of these lines all plants have the same genetic constitution, the high correlation shown here is due to non-genetic influences. That is to say, those plants within a line which have been most favored have developed a thick stem and being more vigorous, because of their favorable environment, have produced the largest amounts of seed. It is obvious that a population composed of the four lines, provided the different lines were not widely different in genetic constitution, would show high positive correlation, whereas, if these same lines were widely different in genetic constitution, their placement might easily minimize or wholly obscure the non-genetic correlation that exists. The manner in which genetic differences in different families may operate to produce correlation or the lack of it in population material is shown in the following correlation tables made from material grown at Lubbock, and while the material in tables 1 and 2

is not distributed according to a normal curve, it serves to show the manner in which families composing a population may take different placement on account of their genetic differences and thus influence the correlation.

Table 1.

Correlation Number of Seed-Bearing Branches and Weight of Threshed Seed. Population Composed of Lines 654 and 223.
F4—1920 Progeny.

Number of Seed-Bearing Branches.

Weight of Seed in Grams.	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60	63	66	
33		1																				1
38	1																					1
43					1																	1
48	1		1	1	1																	4
53		1	1			2																4
58				1	1	1	1															4
63				1		1	1	1	1													5
68				2	1			1														4
73			2	2		4	1	2	1	1						1						14
78				1	1	2	3	2	1							1						11
83					2		3	3	3	1	1					1						14
88					1	1	1	2	1	1	1					1		1	1			11
93				1		1	2	4	2	2	1			1	2	2	1	1	1			21
98							1	2					1		2	2	4					12
103								2	1			1			2	1	2	3	2			14
108								2						1	3	5	9	2	3			25
113														2	1	1	3	6	3	1		17
118															1	3		3	3		1	11
123							1									2	1	3				7
128																2	1			1		4
133															1	1	2	1	1			6
138																			2			2
143																		1				1
148																	1					1
153															1							1
158																						
163																						
168																						
173																						
178																						
183																						
188																						
193																						
198																			1			1
	2	2	4	9	8	12	14	21	10	5	3	1	1	4	13	23	24	21	17	2	1	197

r =.770 ±.019

Table 2.

Correlation Number of Seed-Bearing Branches and Weight of Threshed Seed.
Composition Composed of Lines 567 and 223.
F4—1920 Progeny.

Number of Seed-Bearing Branches.

Weight Seed in Grams.	6	9	12	15	18	21	27	30	33	36	39	42	45	48	51	54	57	60	63	66	66	
33																						
38																						
43																						
48																				1		1
53																		1				1
58																1						1
63															1			1				2
68															1	1						2
73															4	3		1			1	9
78															2	2	1	2				7
83															1	3	2	1				7
88													1		3	2	2	2		1		11
93													1	3	2	2	8	1		1		18
98														3	2	5	4	2	4			20
103														2	3	5	6	4	3			23
108													1	3	6	11	4	4	1	1		31
113													2	1	1	4	8	7	1			24
118														1	3	2	4	6		1		17
123															1	2	4	1		2		10
128															2	4	1		1	1		9
133														1	1	2	2	1	1			8
138																	1	2				3
143															1		1					2
148																1						1
153														1								1
158																						
163																						
168																						
173																						
178																						
183																						
188																						
193																						
198																						
													5	15	34	50	48	36	11	8	1	208

r = .028 ± .046

It is seen here that the population composed of Lines 654 and 223 shows a high positive correlation between number of seed branches and weight of seed, whereas the population composed of Lines 567 and 223 shows no correlation between number of seed branches and weight of seed. Similarly the substitution of other families genetically different may result in a placement of the material so as to show negative correlation.

This material illustrates the manner in which families within a population may take a certain placement because of their genetic constitution and thereby alter the correlation. The correlation in population material may, in fact, be completely reversed if other genetic combinations are present.

Population material has been used to some extent to indicate the intensity of inheritance in a single character between parent and progeny. Here, as in other cases of population material, the coefficient is not necessarily a reliable index to inheritance, the chief difficulty being that in the classification of the parent material its phenotypic rather than its genotypic nature is usually considered. This fact, coupled with the fact that some parents may be homozygous and some heterozygous for the character under consideration, may cause the inheritance to be obscured. This point is emphasized by the work of

Pearl and Surface[15] in breeding for egg production, in which it was shown that where mass selection was practiced no correlation was found between egg production of hens and their daughters; whereas when genotypic selection was practiced and genotypic classes obtained, correlation existed.

The manner in which the correlation in a regression table may vary according to population is illustrated in Tables 3, 4, 5, 6 and 7, showing the correlation between the number of seed-bearing branches in kafir, parent and progeny, when different populations are used. Tables 3, 5, 6 and 7 are based upon selected populations obtained by grouping different families out of an original population from 80 parent heads. Table 3, composed of two radically different families, shows that the material is not distributed according to a normal curve. It is presented, however, as an illustration of the placement of material in regression table where inheritance is present.

Table 3.

Showing Correlation of Seed-Bearing Branches in Parent and Progeny in Population Materi Lines 654 and 223.

Number Seed Branches Progeny, 1921.

Number Seed Branches, Parent 1920.	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60	63	66	69	
3	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
6	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
9	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
12	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
15	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
18	..	..	..	..	..	..	1	..	6	1	2	..	..	..	..	..	..	..	..	..	..	..	..	10
21	..	..	..	1	1	1	7	3	2	4	1	..	..	..	..	..	..	..	..	..	..	..	..	20
24	..	..	..	..	..	2	2	4	1	1	..	..	..	..	..	..	..	..	..	..	..	..	..	10
27	..	..	..	..	3	3	2	5	4	3	..	..	..	..	..	..	..	..	..	..	..	..	..	20
30	..	..	..	..	..	1	3	3	6	5	1	1	..	..	..	..	..	..	..	..	..	..	..	20
33	..	..	..	..	..	1	1	5	2	1	..	..	..	..	..	..	..	..	..	..	..	..	..	10
36	..	..	..	..	..	..	1	4	2	1	1	1	..	..	..	..	..	..	..	..	..	..	..	10
39	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
42	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
45	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
48	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
51	..	..	..	..	..	..	..	..	..	..	..	..	..	1	2	6	13	7	8	3	..	..	..	40
54	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1	1	2	2	3	1	..	..	..	10
57	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	6	12	7	3	1	1	..	..	30
60	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	2	3	2	3	..	..	..	..	10
63	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
66	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	3	4	1	1	..	1	..	10
69	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
	..	..	..	1	4	8	17	24	23	16	5	2	..	1	3	15	33	22	18	6	1	1	..	200

$r = .907 \pm .084$

[15]Data of Pearl and Surface, quoted in "Genetics in Relation to Agriculture," by E. B. Babcock and R. E. Clausen, pages 457-458.

Table 4.

Showing Correlation of Seed-Bearing Branches in Parent and Progeny in Population Material.
80 Families.

Progeny Number Seed Branches (1917)

Parent Number Seed Branches (1916)	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60	63	66	69	72	75	78	81	84	
36																									
39		1	1	1	3	4	2		2				1	1	3	1									20
42											3	4	4	2	6	1	1				1				22
45									1	1	2	5	12	11	12	10	6	1							61
48									1		3	8	17	21	21	14	9		4						98
51											6	2	30	33	34	18	12	5	2	1		1		1	145
54											2	10	14	15	22	12	12	3	3	2					95
57													4	6	9	4	2	2	1						28
60											3	16	26	43	28	27	17	11	3	4	1	1			180
63										1		9	9	7	18	15	4	2				1			66
66									1		5	8	12	22	19	12	4	8	1						92
69											3	8	16	15	25	20	9	7	2		1				106
72												4	5	8	6	4	2	1			1				31
75											1	6	6	8	19	11	9	6	1		1				68
78										1		6	13	12	12	8	1	1							54
	·	1	1	1	3	4	2	0	5	3	28	86	169	204	234	157	88	47	17	7	5	3	0	1	1066

r =.140 ±.020

Table 5.

Showing Correlation of Seed-Bearing Branches in Parent and Progeny in Population Material.
79 Families.

Progeny Number Seed Branches (1917)

Parent Number Seed Branches (1916)	39	42	45	48	51	54	57	60	63	66	69	72	75	78	81	84	
39					1	1	3	1									6
42			3	4	4	2	6	1	1				1				22
45	1	1	2	5	12	11	12	10	6	1							61
48	1		3	8	17	21	21	14	9		4						98
51			6	2	30	33	34	18	12	5	2	1		1		1	145
54			2	10	14	15	22	12	12	3	3	2					95
57					4	6	9	4	2	2	1						28
60			3	16	26	43	28	27	17	11	3	4	1	1			180
63		1		9	9	7	18	15	4	2				1			66
66	1		5	8	12	22	19	12	4	8	1						92
69			3	8	16	15	25	20	9	7	2		1				106
72				4	5	8	6	4	2	1			1				31
75			1	6	6	8	19	11	9	6	1		1				68
78		1		6	13	12	12	8	1	1							54
81																	
84																	
	3	3	28	86	169	204	234	157	88	47	17	7	5	3	0	1	1052

r =.037 ±.020

Table 6.

Showing Correlation of Seed-Bearing Branches in Parent and Progeny in Population Material.
16 Families.

Progeny Number Seed Branches (1917)

Parent Number Seed Branches (1916)	39	42	45	48	51	54	57	60	63	66	69	72	75	78	81	84	87	90	
39																			
42																			
45																			
48																			
51																			
54						3	4	3	3	1									14
57																			
60				1	2	2	1	3	2		1	2							14
63				1	2	2	1	4	2										12
66						2	4	3	1	1									11
69				1	3	8	10	14	3	2	1								42
72				4	5	8	6	4	2	1			1						31
75				1	2	5	4	4	1	3									20
78		1		6	13	12	12	8	1	1									54
81																			
84																			
87																			
90																			
		1		14	27	42	42	43	15	9	2	2	1						198

r =-.281 ±.044

Table 7.

Showing Correlation of Seed-Bearing Branches in Parent and Progeny in Population Material. 16 Families.

Progeny Mean Number Seed Branches (1917)

Parent Number Seed Branches (1916)	39	42	45	48	51	54	57	60	63	66	69	72	75	78	81	84	87	90	
39																			
42																			
45																			
48																			
51																			
54								1											1
57																			
60								1											1
63							1												1
66							1												1
69							2	1											3
72						1	2												3
75							2												2
78						3	1												4
81																			
84																			
87																			
90																			
						4	9	3											16

r = −.742 ± .075

Table 3 shows the correlation existing in a population composed of two families which are radically different with respect to number of seed branches. It is seen that with respect to this particular population there is strong inheritance of number of seed branches. Table 4, being composed of an 80-head population, shows a correlation of .140 ± .020. There can be no doubt that in this population there is inheritance of number of seed branches. Table 5 is a population composed of 79 families obtained by eliminating the one family that has influenced the correlation in Table 4. It is seen that in Table 5 there is no evidence of inheritance. In Table 6, composed of 16 selected families, the correlation has been reversed, showing a coefficient of −.281 ± .044. Table 7 shows this same population of 16 families used by correlating the parent with the progeny means. The result shows a correlation of −.742 ± .075. There can be no doubt that in this population negative correlation exists. The reversal of the correlation in different population material seems dependent upon the extent to which the parent has been influenced by non-genetic factors. For example, family 646, included in the population used in Table 7, came from a parent with a head carrying 69 seed branches and produced progeny with a mean number of 60 seed branches, whereas family 485, also included in Table 7, had a parent head with 78 seed-bearing branches and produced progeny with a mean number of seed branches of 52.38. The failure of the phenotype as a correct measure of the genotype followed by the more nearly correct genotypic placement of the progeny as obtained by the distribution of the progeny or the means of the progeny may result in negative correlation when in fact positive correlation exists. How often this might occur under conditions of random sampling is not known but the material presented shows that it can occur and it is possible that it does occur sufficiently often to justify its consideration as a factor that may influence the correlation.

It is evident that the correlation shown in a regression table based on population material may vary, depending upon the material and the parent classification.

Population material has its limitations as material upon which to base conclusions as to correlation between characters or as to inheritance of a single character between parent and progeny. Accordingly, in the interpretation of correlation data obtained from population material one should not overestimate its value.

THE INTERPRETATION OF PURE-LINE DATA

Pure-line material furnishes a class of data from which one can determine with certainty the measure of correlation existing because of non-genetic influences. That is to say, since all individual plants within a pure line have the same genetic constitution, theoretically at least, the construction of a correlation table of such material would show a placement according to the manner in which these individuals had been affected by environment or other non-genetic influences. We have, for example, correlations in kafir lines grown at Lubbock, Texas, as between weight of green forage and height of plant, as follows:

Line 153: $r = -.212 \pm .056$
Line 192: $r = -.282 \pm .049$
Line 567: $r = -.266 \pm .057$
Line 40: $r = -.257 \pm .057$

We have here, in obtaining the correlation existing in the same season in different lines, a measure of the non-genetic correlation. The correlation in each line is in the same direction and uniformly constant in degree. In this material, non-genetic influences existing have caused a decrease in the weight of green forage as the height of the plant increases, and one may, it seems, accept this as being the normal tendency of non-genetic influences on these two characters.

Again we have a similar case in the correlation between weight of seed and height of plant in different kafir lines, as follows:

Line 153: $r = -.065 \pm .058$
Line 567: $r = -.251 \pm .051$
Line 192: $r = -.233 \pm .050$
Line 40: $r = -.135 \pm .060$

As in the previous case, the direction of the correlation in one line is the same as the direction in other lines. There is a slight variation in the amount of correlation in two cases wherein the coefficients are considered unreliable; nevertheless, there is a consistent tendency within different lines for the taller plants to produce the least seed, which may be taken as a reliable index of the manner in which environment affects these two characters. It is conceivable, of course, that a line might be developed whose characters would react to environmental conditions in a different way to the normal or average line, but no such case has been observed by the writer, and in the event that such a line were established it would not affect the use of several pure lines

as an index to the normal direction of non-genetic influence on correlation.

Pure-line data may also be used to determine the variation from season to season or from condition to condition of the extent of the action of non-genetic influences on the correlation. Love and Leighty[16] have made studies showing the effect of seasonal changes on biometric constants and have used pure-line material for direct comparison from year to year.

Again, one can construct a correlation table to show the relation existing between a single character in the parent and the progeny in line material and determine with certainty whether or not the material in hand is showing genetic variation. The following table is an example of the use of the correlation table in determining whether or not there is variability in a Blackhul White kafir line established at Substation No. 8, Lubbock, Texas.

Table 8.

Showing Use of Correlation Table for Determining Inheritance in Line Material. Line 654.

Number of Seed-Bearing Branches, Progeny.

Number of Seed-Bearing Branches, Parent.	7.5	9.5	11.5	13.5	15.5	17.5	19.5	21.5	23.5	25.5	27.5	29.5	31.5	33.5	35.5	37.5	39.5	41.5	43.5	45.5	47.5	49.5	
9							2	1	2		2	2		1	1								11
10																							
11		1			1			3	2	1				2									10
12	1	1	1	2	1	2		1						1									10
13							1	1	3	2	2	1	1										11
14	1		1		1	3			1	2					1							1	11
15																							
16				1			1		2	3	2		1										10
17			1		2		1	2			2	1							1				10
18																							
19																							
20																							
21																							
22			1				1		1	1	3	2				1							10
23																							
24																							
25																							
26																							
27																							
28																							
29																							
30					1			1	1	3		3	1										10
	2	2	4	3	6	5	6	9	12	12	11	9	3	4	2	1			1			1	93

$r = .032 \pm .070$

It is seen that there is no correlation here and accordingly no genetic variability shown, from which one may conclude that Line 654 with respect to the number of seed branches is quite stable and probably approaches closely to a pure line with respect to this character.

[16]Love, H. H., and Leighty, C. E., "Variation and Correlation of Oats (*Avena sativa*)," Cornell University Agricultural Experiment Station, Ithaca, N. Y., Memoir No. 3, Part I, August, 1914.

THE COMPARATIVE INTERPRETATION OF POPULATION AND PURE-LINE DATA

In view of the nature and causes of correlation, it would seem that the use of both population and line material would afford the most reliable means of making a correct interpretation of data. The pure line gives a reliable measure of the non-genetic influences which may affect characters in relation to each other, and with such knowledge one would seem to be better equipped to detect the action of genetic factors in the population. It is realized that even with a knowledge of the direction and the extent of the effect of non-genetic influences, it will not always be possible to determine the nature of population material in hand, but with this information one may study the correlation table made from population material with greater ease and with a degree of certainty that would not otherwise be possible. For example, take a population composed of two kafir lines, Nos. 654 and 223, showing the distribution of number of seed branches and number of nodes per head, as recorded from material grown at Substation No. 8, Lubbock, Texas.

Table 9.

Showing Population Material in Which the Measure of Non-Genetic Correlation is Known and Illustrating the Use of this Fact in the Interpretation of Population Data.
1921 Data, F5 Material.

Kafir Lines 654 and 223.

Number Nodes per Head.

Number Seed-Bearing Branches.	3	4	5	6	7	8	9	
13	1	1	......	......	......	......	......	2
18	4	6	3	......	......	......	......	13
23	5	24	7	3	......	......	......	39
28	1	17	13	5	......	......	......	36
33	......	2	5	2	......	......	......	9
38	......	1	......	......	......	......	......	1
43	......	......	......	1	1	......	......	2
48	......	......	3	10	9	4	......	26
53	......	......	1	12	19	9	4	45
58	......	......	2	5	8	8	2	25
63	......	......	......	......	......	1	......	1
68	......	......	......	......	......	......	1	1
	11	51	34	38	37	22	7	200

Population, r = .844 ± .013
Line 654, r = .416 ± .055
Line 223, r = .307 ± .061

There is a strong positive correlation here between the number of nodes per head and the number of seed branches. If one has a measure of the correlation due to non-genetic influences, which in this case happens to be + *.416* ± *.055* for Line 654, and + *.307* ± *.061* for Line 223, he is aware of the fact that there is some placement of the families within this table that increases the correlation to more than double the normal effect that one might expect of non-genetic correlations. Hence, one might conclude with certainty that there does exist a genetic correlation between number of nodes per head and number of seed branches. In other words, there are families which possess few nodes and few seed branches and others which possess many nodes and many seed branches. Whether these characters are associated on account of linkage, independent assortment, or because

they are expressions of the same factor, cannot be determined with certainty from the correlation table. In general, however, if the correlation data show that there is always a high positive correlation, it might be inferred that the two characters are probably affected by the same factor. On the other hand, if they are most frequently found to be correlated in a high positive way, but occasionally show a tendency toward reduced correlation, one might infer that linkage with crossing-over existed, whereas if positive correlation is of about the same frequency as negative correlation, independent assortment might be inferred to be the case. The matter of arriving at the cause for genetic correlation is, however, a matter requiring other means of investigation than the use of the correlation table. The point to be emphasized here, however, is the use of the correlation within a pure line as a measure of the non-genetic correlation to be considered in connection with correlation data from populations that the non-genetic influences may be considered in the interpretation of the coefficient of correlation found in populations.

In considering population data, take for example a population made up of the progeny from 80 kafir heads and arranged for the distribution of weight of green forage and the height of plant. Accordingly, we have the following:

Table 10.

Population Composed of Progeny from 80 Heads of Blackhul Kafir.

Weight of Green Forage in Grams.	Height of Plant in Centimeters. 93	108	123	138	153	168	183	198	213	228	243	
150	1	3	1	....	....	....	....	....	....	....	....	5
250	....	17	11	6	3	4	....	....	....	....	....	41
350	....	21	33	40	37	13	....	....	....	....	....	144
450	....	6	53	86	72	30	1	....	....	....	....	248
550	....	1	38	119	44	44	1	1	1	....	....	249
650	....	1	12	72	109	31	....	3	1	2	....	231
750	....	....	1	8	28	3	....	3	....	1	1	45
850	....	....	....	....	1	1	....	....	....	....	....	2
950	...	....	....	....	....	2	....	....	....	....	....	2
	1	49	149	331	294	128	2	7	2	3	1	967

r =.381 ±.017

In the interpretation of this table we have figured the correlation coefficients for these two characters on four different lines, as shown in the following table:

Table 11.

Showing Negative Correlation in Each of Four Different Lines.

Line 153

Weight of Green Forage, Grams.	Height of Plant Cm. 123	138	153	168	183	
150.5	......	......	......	1	......	1
250.5	......	......	3	......	......	3
350.5	......	2	13	4	......	19
450.5	3	12	32	13	......	60
550.5	......	9	27	3	......	39
650.5	......	4	2	1	......	7
750.5	......	1	......	......	......	1
850.5	......	......	......	......	......	
	3	28	77	22		130

r =−.212 ±.056

Table 11—Continued.

Showing Negative Correlation in Each of Four Different Nines.

Line 192

Weight of Green Forage, Grams.	Height of Plant Cm. 123	138	153	168	183	
150.5						
250.5			2			2
350.5		4	6	8		18
450.5		11	15	4		30
550.5	3	28	33	12		76
650.6		16	13			29
750.5		2	1			3
850.5		1				1
	3	62	70	24		159

r =−.282 ±.049

Line 567

Weight of Green Forage, Grams.	Height of Plant, Cm. 123	138	153	168	183	
150.5						
250.5			1			1
350.5			2	1	1	4
450.5		1	11	11	1	24
550.5		1	30	29	2	62
650.5		6	19	13		38
750.5		4	8	5		17
850.5		2	1	1		4
		14	72	60	4	150

r =−.266 ±.051

Line 40.

Weight of Green Forage, Grams	Height of Plant, Cm. 123	138	153	168	183	
150.5						
250.5				2		2
350.5			2	4	1	7
450.5		2	8	5	1	16
550.5	1	12	28	11		52
650.5		7	23	8		38
750.5		1	4	1		6
850.5			1			1
	1	22	66	31	2	122

r =−.257 ±.057

It is seen that we have an average correlation of −.25 as a measure of the correlation of non-genetic influences, whereas in the population material we have a correlation of + .38. It is very evident that in this table made up of population material certain families are tall and produce heavy weights of green forage, whereas other families are dwarf and produce lighter weights of green forage. This is true because the distribution of the material in this population has taken a positive trend in spite of the negative tendencies of non-genetic influences. We may safely conclude, therefore, that there is a positive genetic correlation between height of plant and weight of green forage. And we would further be certain that such families could be isolated and the nature of this genetic correlation determined.

LITERATURE CITED.

Babcock and Clausen, "Genetics in Relation to Agriculture," McGraw-Hill Book Company, Inc., New York, 1918.

Collins, G. N., "Correlated Characters in Maize Breeding," Journal of Agricultural Research, Vol. 6, No. 12, June 19, 1916.

Davenport, E., "Principles of Breeding," Ginn & Co., New York, 1907.

Data of Pearl and Surface, quoted in "Genetics in Relation to Agriculture," by E. B. Babcock and R. E. Clausen, McGraw-Hill Book Company, Inc., New York, 1918.

East, E. M., "Organic Correlations," American Breeders' Association Report, Vol. 4, 1908.

Hayes, H. K., and Garber, R. J., "Breeding Crop Plants," First Edition, McGraw-Hill Book Company, Inc., New York, 1921.

King, W. I., "Elements of Statistical Method," The Macmillan Company, New York, 1921.

Love, H. H., and Leighty, C. E., "Variation and Correlation of Oats (*Avena sativa*)," Cornell University Agricultural Experiment Station, Ithaca, New York, Memoir No. 3, Part I, August 1, 1914.

Webber, H. J., "Correlation of Characters in Plant Breeding," American Breeders' Association Report, Vol. 2, 1906.